Exploring the Users' Preference Pattern of Application Services Between Different Mobile Phone Brands

Hao Jiang, Zhiyi Hu, Xianlong Zhao, Lintao Yang, and Zhian Yang

***Abstract*— User portrait analysis is one of the key points in human behavior analysis. It is important to describe or guess user's characteristics through rational methods in business analysis. In this paper, we use the user details records data set from a mobile operator to analyze the preference of users with different brand phones for different APPs and propose the concept of mobile Internet life personas (MILP) and the latent MILP indexing (LMILPI) model for the analysis of users' MILP. At the same time, we build user portrait analysis framework based on the latent semantic indexing (LSI) theme model, LMILPI model, and association rule mining. On the one hand, we analyze users' preference for APP content when using different mobile brands. On the other hand, we analyze the relationship among mobile brands, user access time, and MILPs to describe users' Internet behavior. Our research shows that there is a difference between users who use different brands of mobile phones: 1) users who use different brand phones have different preferences for different APPs. However, if mobile brand marketing methods or target users are same or similar, the APP preference of these brands will be similar; 2) MILPs are different between users who use different brands of mobile phones, but the MILPs displayed on Android platform are similar even though brands are not same, while the MILP displayed on iPhone is quite different from MILPs on Android; and 3) MILPs' importance will be changed by mobile phone brands and time periods. The analytical framework which we propose can provide commercial solutions such as application recommendations, market strategy formulation, Internet access, and other fields.**

***Index Terms*— APP, association rules, latent semantic indexing (LSI), mobile brand, mobile Internet life personas, user portrait.**

I. Introduction

MOBILE phones have become the necessities of human life and bring kinds of necessary services in social life. In the year of 2016, the sales of smartphones in the world reached 1 billion and 470 million [1]. Mobile applications are the platform for user network life. As of March 2017, the number of Google Play applications has reached 2 million 800 thousand [2], and the number of App Store applications has reached 2 million and 200 thousand [3]. The mobile phones have become users' access to the Internet, so understanding the relationship between mobile phone and users' Internet behavior mode is very important for service providers and mobile operators' recommendation and market strategy formulation.

In this background, mobile phone manufacturers and Internet service providers urgently need to understand the user's requirements and Internet behavior patterns when using different mobile phone brands. On the one hand, by finding out the relationship between APP and brand in user's Internet behavior, mobile manufacturers can improve user experience by preloading APP. On the other hand, service providers can make effective recommendation strategies through the relationship by analyzing the user's transfer behavior between different APPs to attract more people to use their service. Therefore, finding the relationship between mobile phone brand and APP can produce huge economic benefits for both the mobile phone manufacturer and the service provider.

To find out the characteristics of Internet behavior pattern, the traditional method is survey and telephone interview. This method usually consumes more resources, such as time, money, and so on. Another method is the use of user's private information. Previous studies have shown that users' Internet access patterns may be related to many potential factors, such as age, interest, and occupation [4]. However, user's private information is often not accessible to privacy. The user details record (UDR) accurately and anonymously records important information of mobile phone users' Internet activities through cellular network, which becomes the main tool to understand individual behavior characteristics in the era of big data.

However, all types of mobile Internet behavior, such as engaging in Internet activities, doing online shopping, and watching network videos, will generate large amounts of UDR data on the operator side. In such a large scale of data, mining all useful information is obviously hard and not practical. However, we believe that there must be commonalities in user's Internet behavior although the Internet behavior is different and complex. Therefore, digging out commonalities in different Internet behaviors, called typical behavior patterns, from the big data of mobile users is a practical and efficient choice to describe user's mobile Internet behavior.

Manuscript received January 31, 2018; revised August 15, 2018; accepted September 20, 2018. Date of publication November 9, 2018; date of current version December 3, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61371126 and in part by the National Key Research and Development Program under Grant 2017YFB0504103 and Grant 2017YFC0503801. *(Corresponding author: Hao Jiang.)*

H. Jiang and Z. Hu are with the School of Electronic Information, Wuhan University, Wuhan 430072, China (e-mail: jh@whu.edu.cn; huzy94@whu.edu.cn).

X. Zhao is with Beijing Smartchip Microelectronics Company Limited, Beijing 102200, China (e-mail: zhaoxianlongbj@163.com).

L. Yang is with the College of Physical Science and Technology, Central China Normal University, Wuhan 430072, China (e-mail: ltaoyang@mail.ccnu.edu.cn).

Z. Yang is with the National Center for Science and Technology Evaluation, Beijing 102200, China (e-mail: yangzhian@ncste.org).

Digital Object Identifier 10.1109/TCSS.2018.2874466

Therefore, it is our choice to use UDRs to analyze user's Internet activities and typical behavior patterns hidden in different behaviors. To find out the Internet behavior patterns efficiently and to provide business guidance for mobile phone producers and service providers reasonably, we put forward the concept of mobile Internet life personas (MILP). It is a kind of typical behavior patterns which is used to describe user's transfer behavior between different APPs. According to it, we can know the behavior characteristics of users when switching APP, and then we could discover the law of MILP and the relevance between MILP and time, to provide more useful and reasonable business advice. On this basis, we need to understand the following.

1) What is the relevance between different mobile phone brands and different APPs?
2) How to find MILP through UDRs, and what is the homogeneity and heterogeneity of MILP of different mobile phone brands?
3) What is the relationship among the mobile phone brand, the user's Internet time, and the MILP?
4) Therefore, based on the earlier problems, we need an easy and practical method, which can directly display the Internet behavior patterns of users using different brand phones.

We provide a framework for mining mobile users' typical sequential patterns in Internet services, which mainly use thematic models and feature extraction ideas. This framework first analyzes the relationship between mobile phone brand and APP from the point of feature decomposition. Then, the concept of MILP is introduced. It will reveal to us the characteristics of users switching between different APPs. Latent MILP Indexing (LMILPI) thematic model is proposed to discover MILP and weights of MILP. Next, we will find the relationship between MILP and time that users use mobile phones. Through the earlier framework, we can analyze a large amount of UDRs data, transform it into discrete time series, and extract hidden themes and strong rules hidden in the data to design user portraits of different groups of people represented by different kinds of characteristics.

The rest of this paper is set as follows. In Section II, we will summarize the current research status on user typical pattern mining. In Section III, the specific framework and process of the study are explained. The experimental results and discussions are in Section IV. Section V summarizes the work and contribution of this paper and looks forward to the further direction of the future research.

II. Related Work

Many research studies have been done in user portrait analysis, and many effective methods have been put forward to characterize users' habits of Internet.

However, in the traditional methods of analysis, some people are only committed to analyzing the user's Internet habits from the social attributes of the user itself. For example, Luo *et al.* [6] analyze user's behavior through social network models based on the user's social status, and Zhang *et al.* [7] use features of the family place and the workplace to find the interest of the mobile Internet users. These methods reveal a certain correlation between the social attributes of the user and the user's Internet behavior and can effectively analyze user behavior through social attributes. However, most of the data required by these methods belong to personal privacy and are hard to obtain.

Some people try to make use of data generated by users themselves, such as GPS information, browsing logs, and search logs, to analyze users' Internet habits and try to apply them to commercial push. Zhang *et al.* [7] display several characteristic behavior patterns that can guide service providers in application designing, operating, and marketing by the sequence of user web page access. Cole *et al.* [8] analyzed the sequence of user web page access and find that similar patterns of user activity are observed at both the cognitive and page use levels. Godoy [9] uses network document data to design an unsupervised learning method for user portrait. These methods effectively depict users' online habits in content dimension. However, as these methods are only analyzed from the content dimension and ignore other relevant information in the data, such as time information and space information, the information of the data set is not fully used.

Therefore, many people begin to use context to analyze the user's Internet behavior. Trestian *et al.* [11] find that the user's trajectory is associated with the APP usage pattern, and some geographical locations will affect the user's APP usage pattern. Jiang *et al.* [12] use time information in data to predict the time of the Mobile Agent Server and microblog data and find behavior patterns. Han *et al.* [13] analyze the user behavior by exploring the relationship between user interest and social characteristics. However, there are few existing studies that combine the user's social attributes and the data generated by the user to analyze the user's Internet behavior habits.

This paper uses the user's mobile phone brand data and the user's APP data to analyze the homogeneity and heterogeneity of users' Internet habits. In addition, this paper is also an attempt at a joint analysis of the user's attributes and the data produced by the user itself.

III. Models and Methods

A. Data Preprocessing

In this paper, we selected UDR data from a main mobile operator in Beijing, China, including 2 788 384 user entities. The total length of data time is more than 23 days, including the complete 3 weeks. It means that many types of Internet behavior will be recorded in UDR data and provides a guarantee for the richness and accuracy of the experimental results. Before data preprocessing, we made statistics on the proportion of mobile user brand. As shown in Table I, it shows clearly that the top seven brands of mobile phones have accounted for more than 80% of the market share. Therefore, it is meaningful to analyze the relationship between APPs and these brands.

On this basis, 3000 users are selected as samples for each brand concerned, which is used to quantitatively compare the behavior of brands of users.

We have selected some mobile users who have been using mobile phone services for over a year and believe that these

TABLE I
PROPORTION OF MOBILE BRANDS

Mobile Phone Brand	Proportion
Apple	0.348431
Samsung	0.195126
MI	0.111555
Lenovo	0.050958
Huawei	0.046497
NOKIA	0.034173
CoolPad	0.02793
Others	0.18533

TABLE II
DATA STRUCTURE BEFORE PREPROCESSING

User_ID	Start_time	End_time	URL

TABLE III
APP INFORMATION

APP Name	Company	APP Type
QQ	Tecent	Social
Baidu	Baidu	Search
Tieba	Baidu	Social
Zhidao	Baidu	Search
Taobao	Alibaba	Shopping
Alipay	Alibaba	Shopping
Tmall	Alibaba	Shopping
Amazon	Amazon	Shopping
Jingdong	Jingdong	Shopping
Meituan	Meituan	Shopping
Weibo	Sina	Portal
Sohu	Sohu	Portal
UC	UC	Portal
Youku	Youku	Video
Iqiyi	Iqiyi	Video
Kugou	Kugou	Music
Kuwo	Kuwo	Music
PPStream	Iqiyi	Video
Letv	Letv	Video

users have more stable network usage behavior and habits, showing a more typical performance. We use the field structure of UDR data as shown in Table II.

In the era of rapid development of smartphone Internet, the competition of APP is more intense. Therefore, users can choose different applications provided by different service providers for the same use of network behavior to achieve the same goal. Faced with tens of thousands of APP, we choose 20 different APPs based on the market value of service providers and download number of App Store to simplify our experiments, as shown in Table III. Through the URL information of UDRs, we can easily get user app usage record.

On this basis, we sorted the APP records of each user according to the start time of using the app and then form single-user discrete time series which shows the sequence of users' Internet behavior. According to discrete time series, user's transfer behavior between different APPs is clear and we could know which APP will be chosen after user leaves the accessing APP. The process of recording from UDR to discrete time series is completed under Apache Spark. It is a distributed computing framework based on Hadoop and has obvious advantages in large-scale data processing. After data preprocessing, we use the field structure of the UDR data as shown in Table IV.

TABLE IV
DATA STRUCTURE AFTER PREPROCESSING

User_ID	Start_time	APP

B. Experimental Framework

In this paper, a correlation analysis scheme between mobile phone brand and APP is proposed. The program is divided into three parts. The first part mainly uses the latent semantic indexing (LSI) thematic model to find the correlation between brand and APP and the potential hidden theme. Meanwhile, we classify the brands of mobile phones based on their characteristics, which is called brand cluster. In the second part, we introduce the concept of MILP and propose an LMILPI model which aims to discover and extract the MILP hidden in the data. In this way, we can find MILP of different brand clusters. In the last part, we would like to find the change of MILP in different time periods, so we try to analyze the existing phenomena by association rule mining to find the rules hidden in the data set. The program is shown in Fig. 1.

C. Semantic Indexing

To find the correlation between brand characteristics and APP features, many researchers have done a lot of work and achieved notable results. In this paper, we try to analyze the correlation between brand characteristics and APP features from the perspective of topic models.

The LSI is one of the classic models in the theme model, which is often used to find the relationship between words through the mass literature. When some words always appear in a document, there might be a relationship between them, in other words, these words can be regarded as semantically related. By the LSI model, we can quickly analyze the homogeneity between words and get the relationship. At the same time, the LSI can also find the homogeneity of the literature and the correlation between the literature and the vocabulary.

From the algorithm level, the LSI is based on the singular value decomposition (SVD) method to get the subject of the text, as shown in Fig. 2.

In this paper, we think that different kinds of APPs are like words which have their own unique "semantics." These unique "semantics" will match the strongly related "articles," that is, the mobile phone users of different brands. From the point of view of characteristic decomposition, the "semantic" is the feature of APP, where different APPs have different characteristics. Different users also have different features. When user groups with different mobile phones are using the

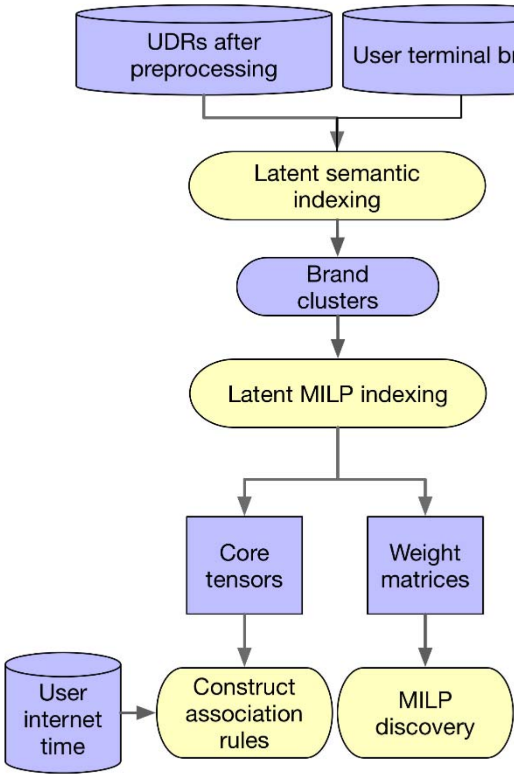


Fig. 1. Photographs of the experimental frame, which shows the relationship among LSI, LMILP, and association rule.

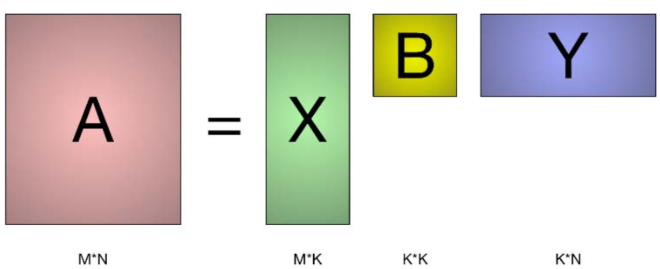


Fig. 2. Schematic of the LSI model. A is a decomposition matrix. The left singular matrix X represents the correlation between the M word and the K semantic class. The right singular matrix Y represents the correlation between the N article and the K theme.

Internet, APP features attract the user group with matching feature. Therefore, we can use the LSI to analyze the correlation between mobile phone brand and APP.

In our experiments, we randomly selected 30 000 users from UDRs to ensure the accuracy and effectiveness of the experiment. In each brand, we count the number of different APPs visits by these users and form Statistica matrix $X_{M\times N}$. M represents the number of brands, N represents the number of APP, and X_{MN} represents the total usage count that users of the brand m use APP n. We processed the matrix X by a logarithmic process and obtained the decomposition matrix A.

We use the SVD decomposition of the matrix A. Based on the size of the eigenvalues, we select the first three column eigenvectors to analyze.

This paper mainly analyzes the relationship between the total usage count and the characteristic values and finds the substitution and complementarity between different APPs. Meanwhile, we divide the brand cluster and the APP cluster based on the hierarchical clustering method and obtain the correlation between the brand and the APP.

D. Latent Mobile Internet Life Persona Indexing

In the current analysis of group behavior, the researchers have found the following phenomenon.

Phenomenon 1: Different users have different behaviors, that is, characteristics of all people are not exactly same [10].

Phenomenon 2: Some users' Internet behaviors are similar and they have same interests, that is, some users have the same or similar characteristics and typical behavior patterns [7].

Phenomenon 3: When users are on the Internet, the current using APP will have an impact on the APP which will be using next time [15].

Based on the above-mentioned phenomenon, we can assume that the user's Internet behavior is determined by the user's own characteristics and the APP used. Therefore, we put forward the LMILPI method based on the above-mentioned conclusion, that is, the LMILPI based on our assumption to find user's characteristics.

Through the LMILPI model, we can analyze the user portrait and find typical behavior patterns hidden in users' behavior. We call it MILP which could describe user's transfer behavior between different APPs and do analysis on users. Next, we will elaborate on the modeling methods and algorithms of the model.

In modeling, based on Phenomenon 1, we think that user U has different MILPs of K, as P_k. Each MILP will have different degrees of influence on user's behavior according to different weights $w(P)$. The greater the weight, the stronger the impact of MILP on users. Based on Phenomenon 3, we believe that different MILPs have different preferences for different APPs, and this difference is affected by the current access of APP. The MILP and the accessing APP, namely, A_t, jointly decide the upcoming APP, namely, A_{t+1}. We define the preference as the access intensity F, whose value is $f(P_k, A_t, A_{t+1})$ and the larger the value, the stronger the access to A_{t+1} preference under *a priori* conditions. Therefore, we put forward the following formula:

$$P(A_t, A_{t+1}) = \sum_k w(P_k) \times f(P_k, A_t, A_{t+1}). \qquad (1)$$

By Phenomenon 2, we can know that some user characteristics are the same or similar, which means that many users may have the same or similar MILP. Therefore, it is meaningful to analyze the MILP shared by users. To analyze the MILP shared by group users, we can transform (1) into tensors, as shown in (2), so we transform the MILP discovery problem of group users into a tensor decomposition problem

$$P(U, A_t, A_{t+1}) = F(P, A_t, A_{t+1}) \times_1 W(P, U). \qquad (2)$$

Among them, $P(U, A_t, A_{t+1})$ is the statistical data, which is the tensor of the probability that the user U has access to A_t at the current moment and will access to A_{t+1} at the next time. $W(P, U)$ refers to the weight of K different MILPs for user U, and $f(P_k, A_t, A_{t+1})$ is the access intensity tensor made up of different MILPs.

At the algorithm level, to decompose the tensor, we combine the nonnegative tensor decomposition called NTF and the high-order SVD (HOSVD) method to deal with the data.

The HOSVD, like SVD, is an SVD method for high-order tensors. It can decompose tensors into three factor matrices and a kernel tensor, and its formula is as follows:

$$\mathbb{R} = \chi \times_1 A \times_2 B \times_3 C. \tag{3}$$

Combined with (2) and (3), we can calculate the tensor decomposition problem based on the HOSVD algorithm, to fully guarantee the nonnegativity of tensors and the independence between features. The formula is as follows:

$$W(P, U) = A \tag{4}$$

$$F(P, A_t, A_{t+1}) = \chi \times_2 B \times_3 C. \tag{5}$$

Based on the above-mentioned modeling method and solution algorithm, we proposed the latent mobile Internet life indexing model to discover the potential MILP of the user.

First, we select 3000 users randomly for each brand cluster generated by the LSI, and we can easily build user transition probability tensor P through their UDR data. Then, we can decompose the probability tensor P of user transfer through the LMILPI model and get the weight of MILP $W(P, U)$ and access intensity tensor $F(P, A_t, A_{t+1})$. To expand access tensor, we can obtain the access K strength matrixes $f(A_t, A_{t+1})$ of different MILPs.

In this paper, we mainly compare the MILP of four different types of brand clusters and analyze the homogeneity and heterogeneity between the MILP of different brand clusters.

E. Association Rule

Association rule mining is a rule-based machine learning algorithm, which can find interesting relationships in large databases. Its purpose is to use some metrics to identify strong rules that exist in a database. Association rules mining is used for knowledge discovery, not prediction. In this paper, we apply the association rule algorithm to find out the rules between MILP extracted from different mobile brand users and users' main Internet time.

First, we analyze the user's online time. We divide 1 day into four periods, of which 0:00–6:00 are time period 1, 6:00–12:00 are time period 2, 12:00–18:00 are time period 3, and 18:00–24:00 are time period 4. We find out the most frequent time of the user's Internet access and mark it as the user's main Internet time T_u.

Then, we calculate the user weight matrix $W(P, U)$ obtained by the LMILPI model and select the MILP with the largest weight as the main MILP of users, which is called P_u.

At the same time, we can get the user's ID and brand information based on UDR data, so that we can get the UDR data of single user U into a set of vectors

$$U \rightarrow (B_u, P_u, T_u). \tag{6}$$

Among them, B_u represents the user U's mobile phone brand cluster, which can be obtained through the LSI model, and P_u refers to the main MILP of the user U, and T_u refers to the user's main Internet access time.

TABLE V
SVD SINGULAR VALUE RESULTS

Singular Value
50.46841
3.164664
2.143987
0.853441
0.636626
0.505439
0.276828

Currently, the problem of seeking correlation among brand cluster, main Internet time and main MILP become the problem of mining association rules for vectors. In the mining of association rules, it is very important to find a reasonable rule. In this paper, we define the rules that need to be analyzed as follows:

$$(U, B_i, T_i) \Rightarrow (U, P_k). \tag{7}$$

In the mining of association rules, two indexes play a decisive role. They are support degree and confidence degree. The support degree represents the probability of the rule appearing in the whole data set, and the confidence degree represents the conditional probability of the rule appearing in the whole data set. In this paper, we find relationships among brand cluster, main time, and main MILP mainly through support and confidence.

IV. Phenomenon Analysis

A. Theme Discovery

We use the LSI theme model to analyze the relationship between the mobile phone brand and the APP. The singular values of the eigenvectors are obtained by the SVD algorithm as shown in Table V. According to the singular value, we can find that the first three items of singular value account for most of the total value of the singular value. Therefore, we try to select the first three eigenvectors to analyze the relationship between the terminal brand and APP.

We found that the eigenvectors corresponding to the maximum singular value have a great relationship with the total amount of users' access. We count the total number of user's hits based on brands and compare it with the first column feature vector of the left singular vector. Surprisingly, the ratio between them is basically constant. In Tables VI and VII, the result shows that the first column of eigenvalues indicates user's hits. We count the number of hits based on APP and compare the result with the first line of the right singular vector, and we get the same conclusion.

Then, we projected the 2-D of the left singular vector and the right singular vector to the plane, as shown in Fig. 3. We can find that different mobile phones show different

TABLE VI
RESULT OF THE LEFT SINGULAR MATRIX ANALYSIS

APP Name	Ratio	APP Name	Ratio
QQ	-133.386	Weibo	-133.83
Baidu	-133.568	Sohu	-133.637
Tieba	-133.519	UC	-133.693
Zhidao	-133.251	Youku	-133.775
Taobao	-133.255	Iqiyi	-132.844
Alipay	-133.665	Kugou	-133.598
Tmall	-132.774	Kuwo	-133.098
Amazon	-133.297	PPStream	-132.945
Jingdong	-133.402	LeTV	-133.365
Meituan	-133.43		

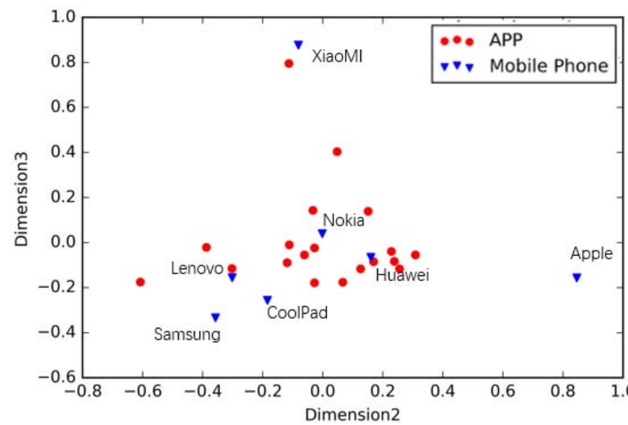


Fig. 3. We selected the singular vector of the second and the third dimension of the left and right singular matrices and then projected them to the 2-D plane as shown in the figure.

TABLE VII
RESULT OF THE RIGHT SINGULAR MATRIX ANALYSIS

Mobile Brand	Ratio	Mobile Brand	Ratio
Apple	-216.15	Huawei	-218.119
Samsung	-217.682	Nokia	-217.848
Mi	-218.677	CoolPad	-218.241
Lenovo	-218.391		

features. Several mobile phones are closer, indicating that they are more similar, and several mobile phones are far away, which means the gap between them is larger. The APP also has the same phenomenon. To make a scientific analysis of it, more specifically, we use the hierarchical clustering algorithm to cluster it, and get several typical "brand clusters" and "APP clusters," as shown in Table VIII.

Through the LSI, the correlation of "brand cluster APP cluster" can be established, and they are roughly divided into the following parts.

1) The potential relationship between mobile phone brands was found. The LSI divides the cell phone into four main categories: "Mi," "Huawei and Nokia," "Samsung and CoolPad and Lenovo," and "Apple." This result is very interesting because it is related to the market positioning of these mobile brands. In 2015, smartphone sales share Samsung, Apple, Huawei, and Mi account for the top four in the world mobile phone market (market share is 24.8%, 17.5%, 8.4%, and 5.6%), and their products positioning and target customers are different. Mi is mainly aimed at users seeking cost-effective and young people. Apple is mainly for the users' who prefer high-quality products. Huawei is mainly used by business people who are always busy. Samsung has products for low-income and middle-income customers who have large free time. Therefore, they are divided into four different clusters. Besides, these main brands "Lenovo" and "CoolPad" are learning the marketing model of "Samsung" and competing with it so the cluster called "Samsung and CoolPad and Lenovo" is build. This phenomenon also appears in the marketing strategy of "Huawei" and "Nokia," so the "Huawei and Nokia" has been set up.
2) From the internal relationship of APP cluster, the competition relationship between applications can be displayed. In the shopping class APP, the "Tmall," "Taobao," "Jingdong," and other shopping applications have a competitive relationship. They are usually not used by the user at the same time, which is different from the traditional impression that the purchase of goods requires more than one shopping software. From the results of the LSI, each user has a customary shopping app and mainly uses it online.
   This phenomenon is also reflected in the video class APP and music class APP. Because of the existence of exclusive copyright, a specific video product cannot be found in two different APPs, and this also forms the competition pattern between them.
3) There is a strong correlation between the mobile phone brand and the APP. For example, users of "Mi" prefer Taobao; users of "Samsung and CoolPad and Lenovo" always use Tmall; "Huawei and Nokia" users use many different shopping APPs. The reason for this phenomenon is the difference in user purchasing power. Users who use Mi mainly focus on cost performance, so they use Taobao to buy cost-effective products. Users of Samsung have begun to focus on the brand of goods, so Tmall, which has more brand stores, is more popular with them. Most of the users of Huawei are business people. They focus on shopping experience and after-sales service. Hence, Jingdong and Amazon are chosen by them.

In the video APP, "Iqiyi" mainly plays what young people like to see, while "Mi" users are mainly young people, so "Mi" is related to "Iqiyi." PPS is the main player in TV play, so "Samsung and CoolPad and Lenovo" users who have free time, will use it. Youku and music have rich content, so business people "Huawei and Nokia" prefer it.

TABLE VIII
CLUSTERING RESULTS OF LSI THEME MODEL

Type	Mobile Brand	APP Information					
		Shopping	Portal	Video	Music	Social	Search
1	Mi	Taobao		Iqiyi			
2	Huawei Nokia	Alipay Meituan Amazon Jingdong	Weibo Sohu UC	Youku LeTV	Kugou	QQ Tieba	Baidu Zhidao
3	Samsung Lenovo CoolPad	Tmall		PPStream	Kuwo		
4	Apple						

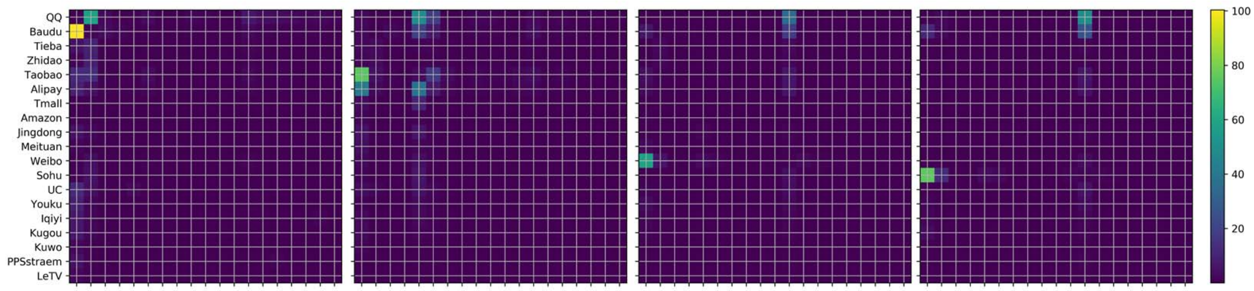


Fig. 4. MILPs of brand cluster 1.

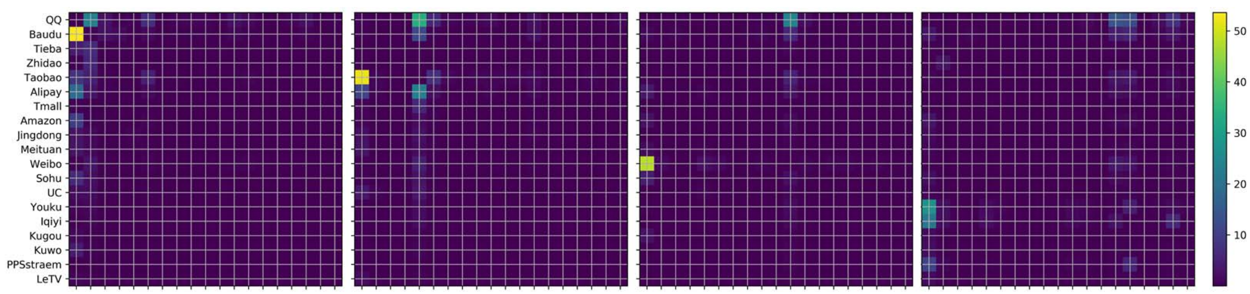


Fig. 5. MILPs of brand cluster 2.

These findings also provide mobile phone manufacturers with proposals for service recommendation and service pre-installation.

The recommendation strategy, which is more consistent with the needs of the user, can achieve a better marketing effect.

B. LIMP Discovery

Based on the LSI, we can make a reasonable classification of the user's mobile phone brand. To balance computation speed and experimental results, we select 3000 users from each brand cluster and use the LMILPI models to find the homogeneity and heterogeneity of the MILP among different brand clusters. To achieve this goal, we show the MILPs of different clusters in the form of a thermal diagram, as shown in the Figs. 4–7.

We have analyzed the results of the decomposition and found that there are similarities and differences between different types of MILP. First, we try to analyze the decomposition results of the first type of brand cluster users in Fig. 4.

The first MILP indicates that the user prefers to choose a "Baidu-QQ" cycle, that is, when users use QQ, they may visit Baidu next and then return to QQ, thus forming a cycle. Meanwhile, when users end up using other APPs, except Baidu and QQ, users are more likely to return to Baidu or QQ and reenter this cycle. This phenomenon is consistent with the results obtained by the LSI model, that is, the two services of Baidu and QQ are divided into one category.

For the second MILP, the social APP such as QQ and the shopping APP, namely, Taobao and Alipay, are more important than others. It is worth to note that, as you can see from the image, the user will have a large probability of using Alipay after using the Taobao. The two APP came from Alibaba. The Taobao is an important shopping platform and Alipay is an electronic payment device. Users will use Alipay to pay money after they choose their favorite products in Taobao, which is also consistent with our reality. Meanwhile, in addition to the strong correlation between the Taobao and Alipay, users still have the preference to switch from QQ to Taobao. This can also help Alibaba to develop advertising strategy, which is to push effectively with Tencent through QQ.

For the third MILP, like the first class of MILP, it forms a "Weibo-QQ" cycle. Weibo is a platform for everyone to

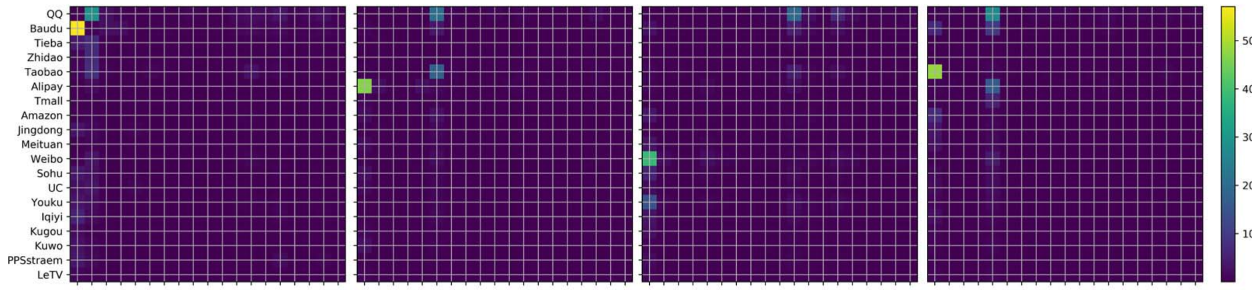


Fig. 6. MILPs of brand cluster 3.

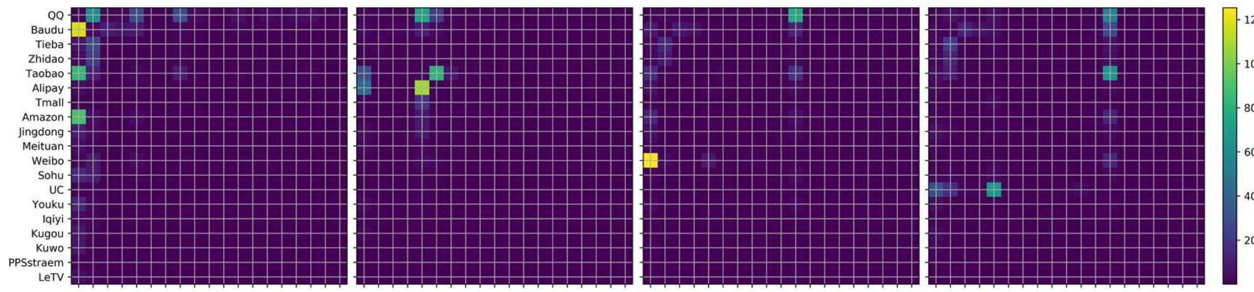


Fig. 7. MILPs of brand cluster 4, and Figs. 4–7 show the MLIPs extracted from the LMILPI model. We use thermodynamic diagrams to depict four typical MLIPs. In the diagram, each row represents the currently used app, A_t, and each column represents the app that will be accessed, called A_{t+1}, and its values show the tendency of the user from A_t to A_{t+1}. In the photograph, the color tends to be yellow, indicating that this tendency is more intense. Tending to blue means a tendency to be weak.

express their views, which likes Twitter. For the fourth MILP, it indicates that users who tend to use SOHU. This is a portal, carrying all kinds of contents and having access to all kinds of services.

According to four MILPs, we can find some interesting results. Baidu, Tecent, and Alibaba are the most important company in China, and our results also show that. Tecent, the owner of QQ, has the most abundant traffic.

When we have completed the internal analysis of the first type of brand cluster users, we try to analyze the similarities and differences between different brand clusters.

We find that mobile phones on Android have similar MILP. The "Mi," "Huawei and Nokia," and "Samsung and CoolPad and Lenovo" all have the "Baidu-QQ" cycle, the "QQ-Taobao-Alipay" cycle and the "Weibo-QQ" cycle. However, there are differences among the fourth MILP. The "Mi" tend to visit SOHU; "Huawei and Nokia" tend to access APP classes such as Youku and Iqiyi; and "Samsung and CoolPad and Lenovo" mainly visit Alipay.

However, there is a significant difference between Apple and the main MILP of these three types of brand clusters. The first MILP of Apple shows that users have a strong tendency to switch from Baidu, Taobao, and Amazon to QQ. The second item has more emphasis on the connection between Taobao and Alipay than the first three. The third item is also "Weibo-QQ." The last item emphasizes the importance of a browser called UC. It can be found that many Apple users use UC browser as an Internet access to access other APP.

Based on the MILP, our analysis can give service providers reliable business advice. For example, it is appropriate for UC to cooperate with Apple and Tecent rather than Mi and Baidu. Therefore, service providers can choose the right mobile phone manufacturers or other service providers to cooperate according to our results.

C. Rules Discovery

We calculate the support degree and confidence degree by the method of association rules and draw them as shown in Fig. 8.

As you can see from Fig. 8, MILP may have a change in different periods of time. For example, MILP1's support degree increases first and then decreases in brand cluster 1, and the confidence degree decreases. In addition, the comparison of the value of support degree between different MILPs will produce different results. In cluster 2, the support degrees of MILP3 are less than MILP4 support in period 3, but in period 4, MILP3 is more than MILP4.

In combination with the phenomenon in the LMILPI, we found that for clusters 1–3, the first three MILPs are basically similar. However, from the point of association rules mining, we can find the difference between the three types of MILP.

First, for brand cluster 1, MILP1 and MILP4 can have a greater impact on users in four time periods. For brand cluster 2, MILP1 is in an important position, while for brand cluster 3, MILP2 plays an important role. This phenomenon shows that even though the MILP decomposed by the LMILPI model is similar, the importance of different MILPs in different clusters is different. This is the complement of LMILPI algorithm.

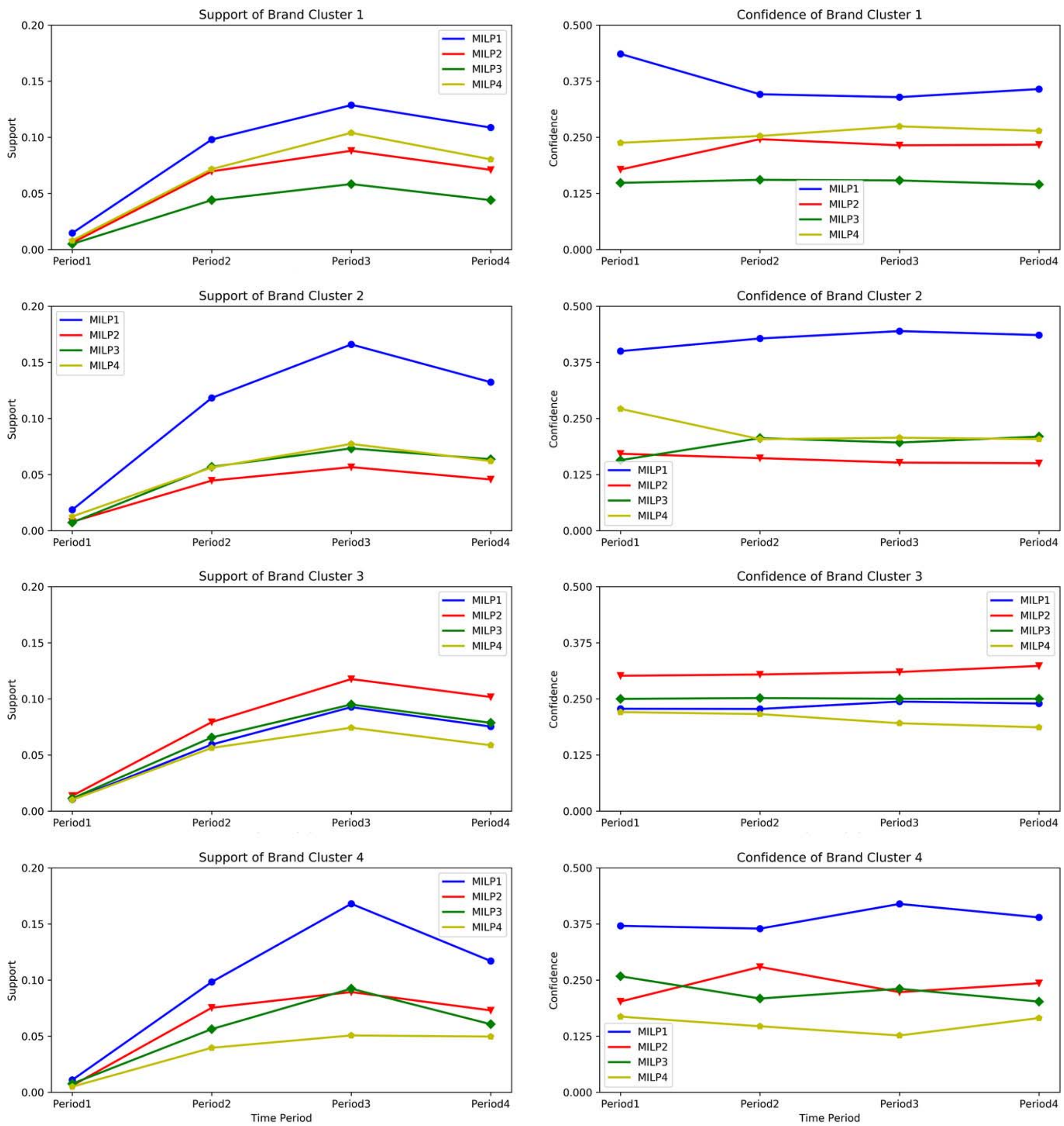


Fig. 8. Four images on the left show the support of different MILPs in the brand cluster 1 to the brand cluster 4. The four graphs on the right show the confidence of different MILPs in the brand cluster 1 to the brand cluster 4.

The analysis of confidence degree is helpful for service providers to make different recommend strategies to brand clusters in different time periods. For example, for brand cluster 4, in periods 1 and 3, the proportion of MILP1 and MILP3 is larger, so we should push Baidu, Taobao, Amazon, and Weibo information when users use QQ. While in period 2, when users use QQ, pushing Baidu, Taobao, and Alipay information is suitable and it is the time that Alipay and Taobao are used at the same moment which means people wants to do online shopping at period 2. Based on these results, users in clusters can get the appropriate APP according to the time information and the user's main MPLI.

Based on support degree, we can analyze the MILP weight matrix of the user and make up for the LMILPI model we proposed. With confidence degree, we can comprehensively analyze the relationship among mobile brand, user's Internet time, and user's main MILP to provide service providers appropriate recommendation strategies based on time.

D. User Portrait Analysis Framework

Based on the LSI, LMILPI, and association rules mining, we propose a framework called user portrait analysis framework to explore typical user behavior. The framework mainly studies the law of user transfer between different APPs and analyzes user behavior based on the mobile phone brand.

In our framework, the LSI will guide the cooperation between service providers and mobile phone manufacturers.

The LMILPI mainly guides the cooperation between service providers. Association rule mining will help service providers know which period is the most appropriate time to recommend. For example, based on LSI, UC can get the information that Apple is the best cooperative business in kinds of mobile manufacturers and UC will know that Weibo and Sohu are its competitors. Based on LMILPI, many people will use UC to do online shopping, so UC should have a close relationship with Taobao. By the way, Alibaba, the company of Taobao, has bought UC in recent years. Based on the association rules, service providers of UC should pull the users' attraction who use iPhone when they hit Taobao, and moreover, afternoon is considered the most suitable time to use. Therefore, service providers can formulate appropriate recommendation strategies through our framework and maximize the effectiveness of recommendations and mobile manufacturers can improve user experience by preloading the APP.

V. Conclusion

The comprehensive development of mobile Internet brings huge business opportunities and profits to mobile phone manufacturers and service providers. However, the expansion of market size is also accompanied by competition. Reasonably finding and analyzing the theme of mobile Internet life using different mobile phone brands can help mobile producers and service providers to seize and increase market share. In this paper, we propose an analysis framework based on the hidden theme model to mining the theme features of mobile Internet users.

The main contributions of this paper include the following.

1) Through the classic theme model LSI, the relationship between different brand mobile phones and different APPs is found.
2) The LMILPI model is proposed and the MILP of users with different mobile phone brands is found and compared based on the LMILPI model.
3) The relationship between the mobile phone brand, the user's main Internet time and the user's main MILP is analyzed.

The above-mentioned work can help mobile phone manufacturers and service providers understand the user's Internet behavior patterns at a certain level. However, in real life, although most users access the Internet through mobile operators, there are still some users who can access the Internet through the wireless network, such as WIFI, which cannot be obtained. Therefore, it cannot fully reflect the portrait of the user.

In addition, this paper uses the HOSVD algorithm in the calculation of the LMILPI model. Although the algorithm provides stable decomposition results, the consumption of computing resources is relatively large, so it is difficult to carry out large-scale data calculation. It is our next step to improve and improve the HOSVD algorithm and reduce its consumption of resources.

In this paper, we extract and compare the users' MILP from different brands of phones based on the feature analysis and topic discovery and do user portrait analysis based on this. This may be ideal. The theme of Internet life is a broad and inaccurate concept. It covers many aspects of human behavior in the network. What we do at present is only a small part of it.

Of course, our current work also provides some ideas for social computing. Other sociological factors, such as user occupation and socioeconomic status, will have an impact on user behavior patterns. Despite the limitations, we believe this research can provide a framework for a new mobile Internet service model.

Acknowledgment

The authors would like to thank Beijing's Advanced Innovation Center for the future network technology for their help.

References

[1] *Worldwide Quarterly Mobile Phone Tracker*, IDC, Framingham, MA, USA, 2017.

[2] Statistia, Hamburg, Germany. *Google Play: Number of Available Apps 2009–2017*. [Online]. Available: https://www.statista.com/statistics/266210/number-of-available-applications-in-the-google-play-store/

[3] Statistia, Hamburg, Germany. *Number of Apps Available in Leading App Stores as of March 2017*. [Online]. Available: https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/

[4] C. Lo, D. Frankowski, and J. Leskovec, "Understanding behaviors that lead to purchasing: A case study of pinterest," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, San Francisco, CA, USA, 2016, pp. 531–540.

[5] J. G. Shanahan and L. Dai, "Large scale distributed data science using apache spark," in *Proc. 21st ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Sydney, NSW, Australia, 2015, pp. 2323–2324.

[6] S. Luo, F. Morone, C. Sarraute, M. Travizano, and H. A. Makse, "Inferring personal economic status from social network location," *Nature Commun.*, vol. 8, May 2017, Art. no. 15227.

[7] X. Zhang, C. Wang, Z. Li, J. Zhu, W. Shi, and Q. Wang, "Exploring the sequential usage patterns of mobile Internet services based on Markov models," *Electron. Commerce Res. Appl.*, vol. 17, pp. 1–11, May/Jun. 2016.

[8] M. J. Cole, C. Hendahewa, N. J. Belkin, and C. Shah, "User activity patterns during information search," *ACM Trans. Inf. Syst.*, vol. 33, no. 1, 2015, Art. no. 1.

[9] D. Godoy and A. Amandi, "Modeling user interests by conceptual clustering," *Inf. Syst.*, vol. 31, nos. 4–5, pp. 247–265, 2006.

[10] A. Girardello and F. Michahelles, "AppAware: Which mobile applications are hot?" in *Proc. 12st Int. Conf. Hum. Comput. Interact. Mobile Devices Services (MobileHCI)*, Lisbon, Portugal, Sep. 2010, pp. 431–434.

[11] I. Trestian, S. Ranjan, A. Kuzmanovic, and A. Nucci, "Measuring serendipity: Connecting people, locations and interests in a mobile 3G network," in *Proc. ACM SIGCOMM*, 2009, pp. 267–279.

[12] M. Jiang, P. Cui, F. Wang, X. Xu, W. Zhu, and S. Yang, "FEMA: Flexible evolutionary multi-faceted analysis for dynamic behavioral pattern discovery," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, New York, NY, USA, 2014, pp. 1186–1195.

[13] X. Han, L. Wang, N. Crespi, S. Park, and Á. Cuevas, "Alike people, alike interests? Inferring interest similarity in online social networks," *Decis. Support Syst.*, vol. 69, pp. 92–106, Jan. 2015.

[14] C. Zhou, H. Jiang, Y. Chen, J. Wu, J. Zhou, and Y. Wu, "TCB: A feature transformation method based central behavior for user interest prediction on mobile big data," *Int. J. Distrib. Sensor Netw.*, vol. 12, no. 10, 2016.

[15] Z.-D. Zhao, Z. Yang, Z. Zhang, T. Zhou, Z.-G. Huang, and Y.-C. Lai, "Emergence of scaling in human-interest dynamics," *Sci. Rep.*, vol. 3, no. 12, 2013, Art. no. 3472.

[16] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.

[17] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by latent semantic analysis," *J. Amer. Soc. Inf. Sci.*, vol. 41, no. 6, pp. 391–407, 1990.

[18] Y. Zhang, M. Chen, S. Mao, L. Hu, and V. Leung, "CAP: Community activity prediction based on big data analysis," *IEEE Netw.*, vol. 28, no. 4, pp. 52–57, Jul. 2014.

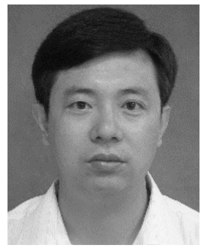
[19] J. Wang *et al.*, "Context-aware role mining for mobile service recommendation," in *Proc. 27th Annu. ACM Symp. Appl. Comput.*, Trento, Italy, 2012, pp. 173–178.
[20] B. Saleh and F. Masseglia, "Discovering frequent behaviors: Time is an essential element of the context," *Knowl. Inf. Syst.*, vol. 28, no. 2, pp. 311–331, 2011.
[21] M. J. Haught, R. Wei, X. Yang, and J. Zhang, "Understanding the psychology of mobile phone use and mobile shopping of the 1990s cohort in China: A lifestyle approach," *Int. J. Online Marketing*, vol. 4, no. 3, pp. 68–84, 2014.
[22] CNNIC, Beijing, China. *The 40th China Statistical Report on Internet Development*. [Online]. Available: http://cnnic.com.cn/IDR/ReportDownloads/
[23] J. L. Zittrain, "Reflections on Internet culture," *J. Vis. Culture*, vol. 13, no. 3, pp. 388–394, 2014.
[24] S. Gálik and S. G. Tolnaiová, "Influence of the Internet on the cognitive abilities of man. Phenomenological and hermeneutical approach," *Commun. Today*, vol. 6, no. 1, pp. 4–25, 2015.
[25] Z. Shen, "Overview: Big data support for urban planning and management in china," in *Big Data Support of Urban Planning and Management*. Cham, Switzerland: Springer, 2018, pp. 1–15.
[26] C. Randler, "Morningness–eveningness and satisfaction with life," *Social Indicators Res.*, vol. 86, no. 2, pp. 297–302, 2008.
[27] J. Shan, F. Joseph, and G. Marta, "Analyzing household lifestyles mobility and activity profiles: A case study of Singapore," in *Proc. Transp. Res. Board 92nd Annu. Meeting*, Washington, DC, USA, Jan. 2013.
[28] S. Luo, F. Morone, C. Sarraute, M. Travizano, and H. A. Makse, "Inferring personal economic status from social network location," *Nature Commun.*, vol. 8, May 2017, Art. no. 15227.
[29] D. Bosomworth, *Mobile Marketing Statistics 2015* (Marketing Intelligence). Leeds, U.K.: Smart Insights, 2015.
[30] J. Park, D.-S. Lee, and M. C. González, "The eigenmode analysis of human motion," *J. Stat. Mech., Theory Exp.*, vol. 2010, no. 11, p. P11021, 2016.
[31] Y. Qiao, X. Zhao, J. Yang, and J. Liu, "Mobile big-data-driven rating framework: Measuring the relationship between human mobility and app usage behavior," *IEEE Netw.*, vol. 30, no. 3, pp. 14–21, May/Jun. 2016.
[32] H. Kawazu, F. Toriumi, M. Takano, K. Wada, and I. Fukuda, "Analytical method of Web user behavior using hidden Markov model," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2016, pp. 2518–2524.

Hao Jiang received the B.Eng. degree in communication engineering and the M.Eng. and Ph.D. degrees in communication and information systems from Wuhan University, Wuhan, China, in 1999, 2001, and 2004, respectively.

From 2004 to 2005, he was a Post-Doctoral Researcher with LIMOS, Clermont-Ferrand, France. He was a Visiting Professor with the University of Calgary, Calgary, AB, Canada; ISIMA, Aubière, France; and Blaise Pascal University, Clermont-Ferran. He is currently a Professor with Wuhan University. He has authored over 60 papers in different journals and conferences. His current research interests include mobile *ad hoc* networks and mobile big data.

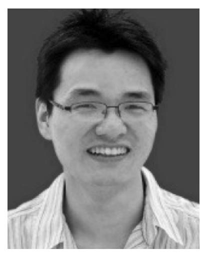
Zhiyi Hu was born in Shaoyang, Hunan, China, in 1994. He received the bachelor's degree from the College of Communications Engineering, Chongqing University, Chongqing, China, in 2016. He is currently pursuing the M.S. degree with the Department of Information and Communication Engineering, School of Electronic Information, Wuhan University, Wuhan, China.

His current research interests include data mining, machine learning, and distributed computing.



Xianlong Zhao received the bachelor's degree from the Hebei University of Technology, Tianjin, China, in 1996, and the master's degree from North China Electric Power University, Beijing, China, in 1999. He is currently pursuing the Ph.D. degree with the School of Electronic Information, Wuhan University, Wuhan, China.

His current research interests include wireless communication networks, data mining and large data analysis, power distribution technology, and power system communications.



Lintao Yang received the M.Eng. and Ph.D. degrees in communication networks from Wuhan University, Wuhan, China, in 2005 and 2011, respectively.

He is currently a Lecturer with the College of Physical Science and Technology, Central China Normal University, Wuhan. His current research interests include complex networks and wireless *ad hoc* networks.

Zhian Yang was born in Xianyang, Shaanxi, China, in 1985. He received a bachelor's degree from the University of Science and Technology of China, Hefei, China, in 2007, and the Ph.D. degree from the Chinese Academy of Sciences, Beijing, China, in 2012.

From 2012 to 2013, he with the 54th Research Institute of China Electronics Technology Group Corporation (CETC54), Shijiazhuang, Hebei, China. From 2013 to 2016, he was involved in search work in China. Since 2016, he has been with the Science and Technology Assessment Center, Ministry of Science and Technology, Beijing, China, where he is involved in the research and technical management of advanced information technology.